

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: October 28, 2013

K.P. Patel
J.M. Medved
R.F. Fernando
B.P. Pithawala
Cisco Systems
April 26, 2013

Service Advertisement using BGP
draft-keyupate-bgp-services-02.txt

Abstract

A variety of services, such as NATs, firewalls, or caches, can be embedded in a service provider network or instantiated in data centers attached to the network. This document proposes extensions to BGP that facilitate discovery of such service instances by interested clients and allows dissemination of service information, such as services capabilities or capacities, throughout the network domain. The proposed extensions allow for optimal routing of requests to service instances that can optimally serve them.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 28, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	3
1.1.	Requirements Language	3
1.2.	Terminology	3
1.3.	Overview	4
2.	Service Advertisement & Discovery	5
2.1.	The Service Advertisement Attribute	5
3.	Distribution of Service Data	6
3.1.	Capability Advertisement for Service AFI	6
3.2.	BGP Service AFI	6
3.3.	NLRI Format for Service AFI	7
3.4.	The Service Attribute	8
4.	Operation	8
4.1.	Announcing BGP Service Discovery Attribute	8
4.2.	BGP Service AFI	8
5.	Service Example	9
6.	IANA Considerations	10
7.	Security Considerations	10
8.	Acknowledgements	10
9.	References	10
9.1.	Normative References	11
9.2.	Informative References	11
	Authors' Addresses	12

1. Introduction

The current BGP specification does NOT provide any generic mechanism to advertise service instances within a BGP enabled network. The current BGP specification also does NOT provide a mechanism to carry opaque service data within a BGP enabled network. This draft defines several extensions to BGP that allows both the advertisement and the discovery of service instances and the distribution of service related data within a BGP network.

Service advertisement and discovery is facilitated by distributing Service Advertiser location information throughout the network. This document proposes a new optional transitive attribute - the Service Advertisement Attribute that carries addresses of BGP Speakers, thereby facilitating the service advertisements and service discovery. Entities interested in receiving service information will peer with one or more service-advertising BGP Speakers and receive service data directly from them.

This draft also defines a new BGP Capability known as BGP Service AFI Capability, a new BGP AFI and its NLRI format to carry opaque Service data within a BGP enabled network. BGP Service AFI leverages existing BGP machinery of message announcements, processing of received messages, loop detection, policy application and other vital protocol framework to transport Service information. Routers enabled with Service AFI are expected to store Service information possibly as an opaque data.

In addition to the service information discovery and dissemination mechanism, Pub/Sub APIs MAY be required for registration of service instances with the network entities. The definition of such Pub/Sub APIs are outside the scope of this document.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",

"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

[1.2.](#) Terminology

This document uses the following terminology:

Service Instance: A Service application running on an end device or on a network device. Service applications are broadly classified into two categories: Network Services and Application Services.

Patel, et al.

Expires October 28, 2013

[Page 3]

Internet-Draft

Service Advertisement using BGP

April 2013

Service consumer: An end device or a network device that is a interested in a particular Service.

Gateway: An network device that is responsible for doing a service conversion across different protocols.

Originator: An end device or a network device originating a particular service.

[1.3.](#) Overview

In today's networks, both the services provided in the network (NATs, Firewalls, etc.) and the service overlays (services provided in enterprise or cloud data centers) are dynamic in nature. New service instances are instantiated on demand, instances that are not sufficiently utilized are deleted and their resources are redeployed; load and traffic patterns change frequently and abruptly; more and more service consumers are mobile. A service consumer (or its proxy, such as a service router) needs to know which services it can reach and their location. To optimize the usage of the underlying network and the service overlay, a service consumer needs to know where to find service instances and what are capacities and capabilities of each service instance.

BGP provides a transport to distribute service information for several reasons:

- o Scalability, performance and ubiquitous deployment in service providers networks

- o It can be used to distribute information within an AS domain or between AS domains
- o Leverage its security
- o Rich policy capabilities allow the operator to control the distribution of service information

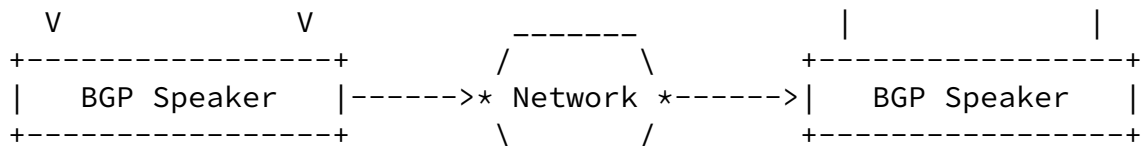
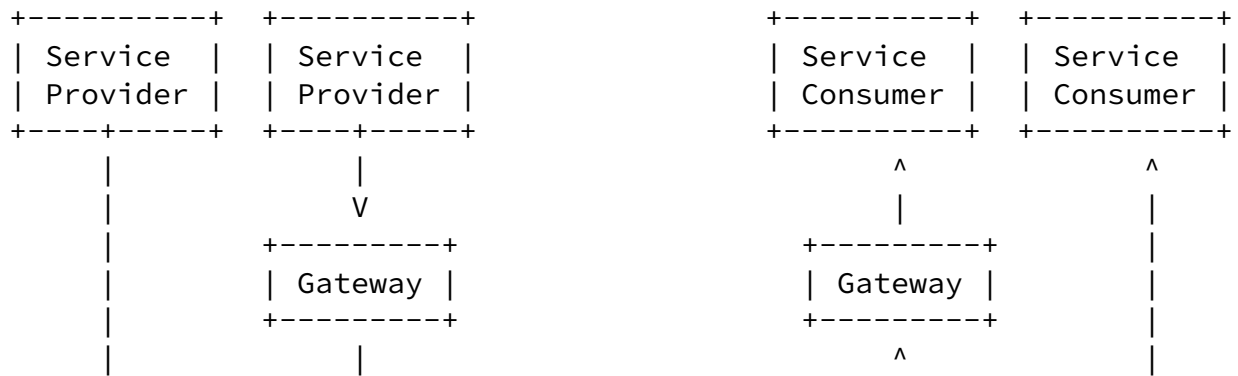


Figure 1: Service discovery and metadata exchange

Figure 1 shows a typical service network where services consumers interact with each other over a network. As shown in the figure, a Service Provider may inject service information into BGP either directly or through a Gateway. BGP then distributes the Service information within its network. Service information may also be distributed across the AS domains. A Service consumer receives Information about available services either directly by listening to BGP updates or through a Gateway. Gateways may provide different APIs to Service Providers/Consumers, for example ReST, XMPP, or Websockets.

A Gateway translates between an application-friendly data format / API on the Service Provider/Consumer Side, and a BGP-specific format on the BGP Speaker side. The BGP-specific format would use BGP protocol (i.e the Gaetway appears as a peer to the BGP Speaker) or an

API that can inject/retrieve data directly from BGP internal databases, such as IRS.

2. Service Advertisement & Discovery

2.1. The Service Advertisement Attribute

The Service Advertisement attribute is a new BGP optional transitive attribute. The attribute type code for the Service Advertisement attribute is to be assigned by IANA. The value field of the Service Advertisement attribute is defined as set of one or more Service Tuples.

A Service Tuple within the Service Advertisement Attribute is defined as follows:

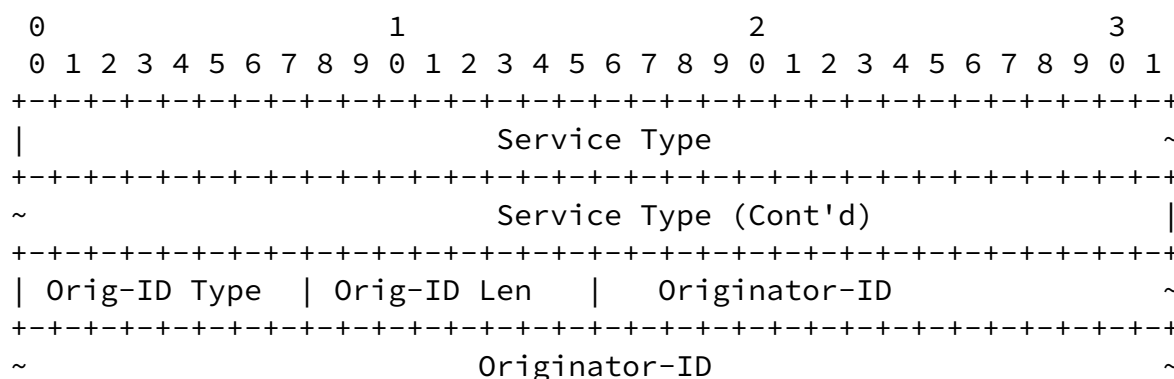
[illegible]

Figure 2: Service Tuple format

The fields are as follows:

Service Type: Eight octets encoding the Service Type. The Service Type code point space will have to be maintained by IANA. Only type 1 is defined in this document. Service ID from 4294967296 onwards are reserved for private use only.

Originator-ID Type: One octet encoding the Originator-ID Type. Type value of 0 indicates an IPv4 address type, Type value of 1 indicates an IPv6 address type, and Type value of 2 indicates an AS number type.

Originator-ID Length: One octet encoding the Originator-ID address length. Originator-ID is not present in the Service Tuple if the Originator-ID Length is 0.

Originator-ID: Optional, contains the IP address of the router Originating the Service Tuple. If the advertising router's IP address is not present, then the Originator of the Service Advertisement attribute is the first AS aka rightmost AS in the final segment of an AS_PATH attribute if that segment is of type AS_SEQUENCE, or the distinguished value "NONE" if the final segment of the AS_PATH attribute is of any type other than AS_SEQUENCE. Originator ID length of 4 octets indicate an IPv4 Address and 16 octets indicates an IPv6 address

The service attribute applies to all prefixes carried in an UPDATE message - either in IPv4 NLRI or in MP_REACH/MP_UNREACH attributes.

[3.](#) Distribution of Service Data

[3.1.](#) Capability Advertisement for Service AFI

The BGP Service AFI Capability is a new BGP capability, that can be used by a BGP speaker to indicate its support for BGP Service AFI Advertisements. A BGP speaker willing to exchange this capability MUST advertise this information in the OPEN message using BGP Capability code 1 [[RFC4760](#)]. The AFI value is set to BGP Service AFI value to be assigned by IANA. The SAFI value is set to UNICAST value.

[3.2.](#) BGP Service AFI

This document introduces a new AFI known as a BGP Service AFI with a value to be assigned by IANA. The purpose of this AFI would be to exchange Service specific information within a BGP network. The SAFI value is set to UNICAST value as defined in IANA SAFI registry.

[3.3.](#) NLRI Format for Service AFI

BGP Service AFI NLRI is advertised in BGP UPDATE messages using the

MP_REACH_NLRI and MP_UNREACH_NLRI attributes [[RFC4760](#)]. The AFI used to identify this NLRI is BGP Service AFI with a value to be assigned by IANA. The SAFI value is set to UNICAST value as defined in IANA SAFI registry.

The Next Hop field of MP_REACH_NLRI attribute shall be interpreted as an IPv4 address whenever the length of the Next Hop address is 4 octets, and as an IPv6 address whenever the length of the Next Hop address is 16 octets.

The NLRI field in the MP_REACH_NLRI and MP_UNREACH_NLRI is a prefix with the maximum length of 276 octets. The NLRI field of the MP_REACH_NLRI and MP_UNREACH_NLRI is encoded as a 1-octet NLRI length field in bytes followed by a variable-length NLRI value. The NLRI length is expressed in octets.

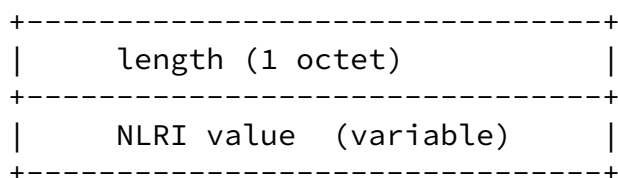


Figure 3: Service NLRI

The Length field indicates the length of NLRI value in octets.

The NLRI Format is as follows:

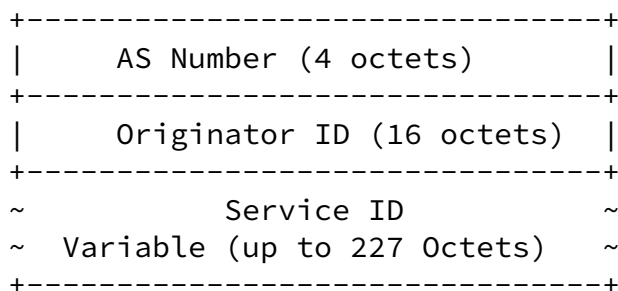


Figure 4: Service NLRI format

The fields in the Service NLRI are as follows:

System originating the prefix. The AS Number value will be the value assigned by IANA.

Originator ID: 16 Octets encoding the Originator ID value.

Service ID: Upto 227 octets encoding the Service ID value.

[3.4.](#) The Service Attribute

The Service attribute is a new BGP optional transitive attribute. The attribute type code for the Service attribute is to be assigned by IANA. The Service attribute carries BGP Service AFI NLRI specific information. The value field of the Service attribute contains service specific data opaque to BGP.

[4.](#) Operation

[4.1.](#) Announcing BGP Service Discovery Attribute

The Service Discovery attribute is used to announce/withdraw new Service instances within a network. The Service Discovery attribute is a new optional transitive attribute that can be applied to any AFI/SAFI within BGP; thereby facilitating service instance discovery on an existing network. The attribute is not related to the AFI/SAFI to which it is attached, and it does not impact the BGP selection algorithm. The AFI/SAFI is used merely as a distribution mechanism for Service Advertiser information carried in the attribute.

A BGP Speaker MUST aggregate all the Service attribute Tuple information received from all the paths (including the non-bestpaths) for a given NLRI and announce the aggregated Service information along with the bestpath in case where all the BGP paths [ADDPATHS] are not being announced. A BGP Speaker MUST announce BGP Service attribute along with its own path whenever all the paths [ADDPATHS] are announced. These rules ensure that Service Discovery attributes always get announced within a given network.

Each Service Announcement attribute tuple consist of an optional Originator-ID value. If this value is present, then a Service Tuple information is considered as incomplete if it does not contain a valid Originator-ID address or an Origin AS number. In such conditions, the received Service Tuple information maybe be ignored with an appropriate error logging.

[4.2.](#) BGP Service AFI

BGP Service AFI is a new AFI used to carry Opaque Service information. The opaque Service data is carried within BGP using the BGP Service Attribute that is attached to a BGP Service AFI NLRIs. BGP Speakers configured to exchange data over the BGP Service AFI are required to store and forward the opaque Service level information. A standard BGP refreshing mechanisms [[RFC2918](#)] could be used to refresh the information hop-by-hop whenever required. The scope of announcing opaque Service data could be control using BGP Communities defined in [[RFC1998](#)] or using BGP Extended Communities defined in [[RFC4360](#)].

BGP Service AFI MAY use multisession [[I-D.ietf-idr-bgp-multisession](#)] to have a separate session to transport its opaque data and thereby not affect the performance of existing routing AFI/SAFIs within BGP.

5. Service Example

It is assumed that service-specific attributes and NLRIs will be defined in their own documents. Each service needs to advertise its reachability, capacities and capabilities. This section provides an example of a service - a simple caching service.

It is assumed that the Service ID will be allocated by IANA for the simple caching service.

The simple caching service will only advertise service availability during service discovery using the Service Advertisement Attribute ([Section 2.1](#)). A simple cache service instance is identified by its IPv4 and IPv6 addresses. The Service ID in the Service NLRI (Figure 3) is defined as follows:

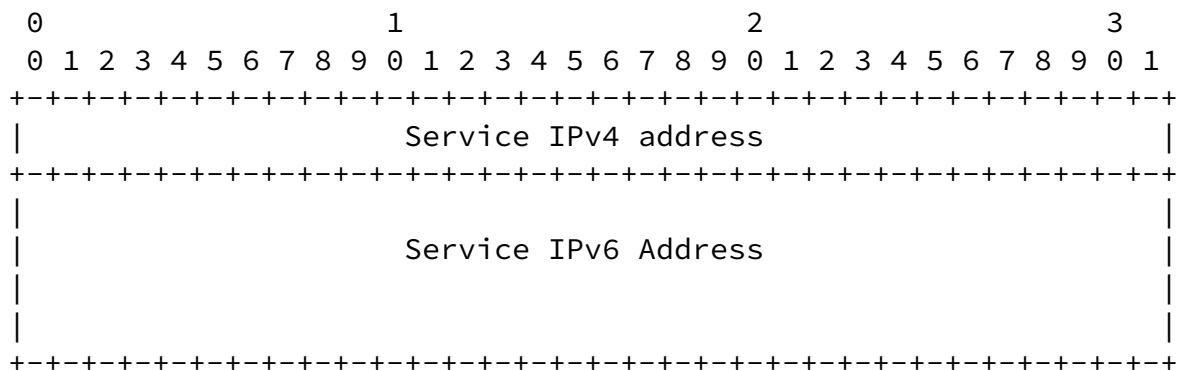


Figure 5: Service ID for simple cache service

A simple cache service instance advertises its capacities and capabilities as TLVs in the service Attribute ([Section 3.4](#)). The TLV

data is opaque to the BGP speaker. The TLVs are as follows:

Load: Identifies the current load of the cache; the bigger the load, the less preferred is this instance of the service. Load is an integer in the range 0-100. which shows the percentage of the capacity of the service instance is currently being used by clients.

Supported media types: This TLVs lists all media types that can be served from this cache instance. There is a sub-TLV for each media type.

Protocols: This TLVs lists all protocol types that can be used to serve content from this cache instance. There is a sub-TLV for each protocol type.

Simple cache service instance advertisements are injected into BGP by simple cache instances through a gateway or a proxy providing an XMPP or RESTful API. Simple cache service instance advertisements are received by a service router, which directs client requests to the most appropriate instance of the cache.

6. IANA Considerations

This document requests a code point from the registry of Address Family Numbers.

This document requests a code point from the BGP Path Attributes registry.

This document requests creation of a new registry for service type codepoints. Allocations within the registry will require documentation of the proposed use of the allocated value and approval by the Designated Expert assigned by the IESG (see [[RFC5226](#)]).

Note to RFC Editor: this section may be removed on publication as an RFC.

7. Security Considerations

This extension to BGP does not change the underlying security issues

inherent in the existing [[RFC4724](#)] and [[RFC4271](#)]

[8.](#) Acknowledgements

The authors would like to thank Dave Ward and Stefano Previdi for the review and comments.

[9.](#) References

Patel, et al. Expires October 28, 2013 [Page 10]

Internet-Draft Service Advertisement using BGP April 2013

[9.1.](#) Normative References

- [I-D.ietf-idr-bgp-multisession]
Scudder, J., Appanna, C., and I. Varlashkin, "Multisession BGP", [draft-ietf-idr-bgp-multisession-07](#) (work in progress), September 2012.
- [RFC1998] Chen, E. and T. Bates, "An Application of the BGP Community Attribute in Multi-home Routing", [RFC 1998](#), August 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2918] Chen, E., "Route Refresh Capability for BGP-4", [RFC 2918](#), September 2000.
- [RFC3392] Chandra, R. and J. Scudder, "Capabilities Advertisement with BGP-4", [RFC 3392](#), November 2002.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), February 2006.
- [RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y. Rekhter, "Graceful Restart Mechanism for BGP", [RFC 4724](#), January 2007.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", [RFC 4760](#), January

2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008.

[9.2.](#) Informative References

[RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", [RFC 4486](#), April 2006.

Patel, et al. Expires October 28, 2013 [Page 11]

Internet-Draft Service Advertisement using BGP April 2013

Authors' Addresses

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: keyupate@cisco.com

Jan Medved
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: jmedved@cisco.com

Rex Fernando
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: bpithaw@cisco.com

Burjiz Pithawala
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: bpithaw@cisco.com